

#16
BP
7-19-02RECEIVED
MAR 14 2002
TECH CENTER 1600/2900

ENTERED

1632

RAW SEQUENCE LISTING
PATENT APPLICATION: US/09/950,051DATE: 03/04/2002
TIME: 15:03:28Input Set : A:\34950209.app
Output Set: N:\CRF3\03042002\I950051.raw

3 <110> APPLICANT: YERAMIAN, EDOUARD
 5 <120> TITLE OF INVENTION: GENES AND THE PHYSICS OF THE DNA DOUBLE HELIX.
 6 FORMULATION OF A PHYSICS-BASED GENE IDENTIFICATION
 7 (PBGI) METHOD: AB INITIO IDENTIFICATION OF GENES IN
 8 EUKARYOTIC GENOMES
 10 <130> FILE REFERENCE: 03495-0209-00000
 12 <140> CURRENT APPLICATION NUMBER: 09/950,051
 13 <141> CURRENT FILING DATE: 2001-09-12
 15 <150> PRIOR APPLICATION NUMBER: 60/232,146
 16 <151> PRIOR FILING DATE: 2000-09-13
 18 <160> NUMBER OF SEQ ID NOS: 9
 20 <170> SOFTWARE: PatentIn Ver. 2.1
 22 <210> SEQ ID NO: 1
 23 <211> LENGTH: 213
 24 <212> TYPE: DNA
 25 <213> ORGANISM: Plasmodium falciparum
 27 <400> SEQUENCE: 1
 28 atgtgcatac atgttacgtt taatttttat ttgaagata atgatttttag tgcgttgaaa 60
 29 gtttaaggatg aagaaattgt ttctaagaaa aataatttct cttttctgct tcttagcaat 120
 30 gattcaaatt ctgtaacaaa aaagtacata gttgatttga cttactaga taatattata 180
 31 gaatccgtaa gaaataaaaag aaatataaaa aga 213
 34 <210> SEQ ID NO: 2
 35 <211> LENGTH: 212
 36 <212> TYPE: DNA
 37 <213> ORGANISM: Plasmodium falciparum
 39 <400> SEQUENCE: 2
 40 agtatatttt ttgaacatc aaattttcgc atcgttgagg ctccccaggt gcgttgaaa 60
 41 ttaaggatga agaaattgtt tctaagaaaa ataatttctc cttttctgct ctttagcaat 120
 42 attcaaattc tgtaacaaaa aagtacatag ttgatttgac cttactagat aatattatag 180
 43 aatccgaaac caaatacaat ttgcttctg tg 212
 46 <210> SEQ ID NO: 3
 47 <211> LENGTH: 225
 48 <212> TYPE: DNA
 49 <213> ORGANISM: Plasmodium falciparum
 51 <400> SEQUENCE: 3
 52 atgtatagac gcatacatat tattacattt gtaacgatca atcttttttt cttattatcc 60
 53 ctatcccaca gatcatcatga tagcgtccag aatttcttga aggaagaaaa aaataactct 120
 54 gataagttac aagatgatat agatgaggat gaggaaaaat attttgacga ggaaatttta 180
 55 aggggaagcca aaaaaaaaaag tgaagaatat gataaagacg atgaa 225
 58 <210> SEQ ID NO: 4
 59 <211> LENGTH: 225
 60 <212> TYPE: DNA
 61 <213> ORGANISM: Plasmodium falciparum

RECEIVED
MAR 14 2002
TECH CENTER 1600/2900

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/950,051

DATE: 03/04/2002

TIME: 15:03:28

Input Set : A:\34950209.app

Output Set : N:\CRF3\03042002\I950051.raw

63 <400> SEQUENCE: 4

```

64 aaagaaaaac gcatacatat tattacattt gtaacgatca atcttttttt cttattatcc 60
65 ctatcccaca gatatcatga tagcgtccag aatttcttga aggaagaaaa aaataactct 120
66 gataagttac aagatgatat agatgaggat gaggaaaaaat attttgacga ggaaatttta 180
67 agggaagcca aaaaaaaaaat aacattcttt tcacacgatt ttgtt 225

```

70 <210> SEQ ID NO: 5

71 <211> LENGTH: 600

72 <212> TYPE: DNA

73 <213> ORGANISM: Homo sapiens

75 <400> SEQUENCE: 5

```

76 ataaacattc tttagtccac acatagataa ataaataagg aagcaaatag acacacagaa 60
77 gagcgggaca gctcctcctc ccgggagaaat ttcaattagt aagtgtggaa ggaacaaggc 120
78 agggaggaga atcctcaaca gagccccaca gggaccgtgc gggcgaggcc cccggagggg 180
79 caccagcact gccgggcaaa cgctgggca gagcggggac agctgccaaag tctcagacat 240
80 gaccaattac agagggaaac ggcggcaccg cgagggatgg gccgcggccg tgtcacctcc 300
81 atgccccacg cacactgctc ctgtgggatt cctcccccaa cacgatgcc actctgacca 360
82 cgaggaaacc tcaagcaagt ccacgtggag gggcattcta caaacacccc aaccgggtcaa 420
83 ggtcgctgag gccaaaggaga gattgggcaa ccgtcacaaa ccagagaagc cgaggagagc 480
84 ttccagccaa cgccatgtgg ggtcctgagc aggaccacc ggaagtgtgt gcagctgcct 540
85 aaagaccgtc ctggctgaga agaaacagag cagcgtctgt ttctcagagc tgggaaccga 600

```

88 <210> SEQ ID NO: 6

89 <211> LENGTH: 1980

90 <212> TYPE: DNA

91 <213> ORGANISM: Homo sapiens

93 <400> SEQUENCE: 6

```

94 acctgatct cagacttctg gtttccaaaa ccatgagaca cggaatttct gttgtgtgac 60
95 cagccagttt gtggtactgt ttgtcatggc agcccaagga aaagaataca ttacagcata 120
96 caaacatga ctacattat ctttacttag aaccctctct cctaagcttt 180
97 caatcacaga ggcacatgat cttgttcagc agcctagaaa accaaggccc agcggagcca 240
98 ccgtaggca cccactcccc atagcctggc acacacacac ggcagagcca cccacaggca 300
99 cccactcctc atagtccagc acacacacgg cagagccacc cgcaggcacc cactccccat 360
100 agcccgccac acacgtggac catgccaccc tccacgtgcg cctggggagc aaagcagcac 420
101 agcctgaact gccctcagc tcttctctct gagtctaaaa cagcacatg cgtcccaggc 480
102 caattccaag ttttgtaaac tgagcaacag ctcttgggaa acaaaaaacac agctactgtt 540
103 tattctctct gagctggctg tacaccccaa caaggaaggg agggcttgtg gagcctctct 600
104 tctggacaac atgcaccaag gaggagtata aaagccccc aaaccgagc acctactca 660
105 ctgctcacc cactccctcc catctcccc agctcaaccc ccagcacagc agcatccacc 720
106 atgtccgtct gctccagcga cctgagctac agcagccgcy tctgccttcc tggttcctgt 780
107 gactcttctg ccgactcctg gcaggtggac gactgccagc agagctgctg cgagccccc 840
108 tgetgcgccc ccagctgctg cgccccggcc cctgctga gctgtgtctg caccctagt 900
109 agcgtgtgt ccagcccctg ctgcccagtg acctgcgagc ccagcccctg ccaatcaggc 960
110 tgcaccagct cctgcacgcc ctgctgctgc cagcagtcta gctgccagct ggcttgcctg 1020
111 gctcctccc cctgccagca ggctgctgc gtgccctct gctgcaagac tgtctgctgc 1080
112 aagcctgtgt gctgtgtgcc cgtctgctgt ggggattctt catgctgcca gcagtctagc 1140
113 tgccagtcag cttgctgcac ctctccccc tgccagcagg cctgctgtgt gcccatctgc 1200
114 tgcaagcctg tctgctctgg gatttctct tctgctgcc agcagtctag ctgtgtgagc 1260
115 tgtgtgtcca gccctgctg ccaggcgggc tgtgagccca gccctgcca atcaggctgc 1320
116 atcagctcct gcagccctc gtgctgccag cagtctagct gccagccggc ttgctgcacc 1380
117 tctctctct gccagcaggc ctgctgcctg ccgctctgct gcaagactgt ctgctgcaag 1440

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/950,051

DATE: 03/04/2002

TIME: 15:03:28

Input Set : A:\34950209.app

Output Set: N:\CRF3\03042002\I950051.raw

```

118 cctgtgtgct ctgaggattc ctcttcatgc tgccagcagt ctagctgcca gccggcttgc 1500
119 tgcaacctct ctccctgcca gcaggcttgc tgtgtgcctg tctgtgcaa gcctgtgtgc 1560
120 tgcaagcctg tcggctctgt gcccatctgc tctggggctt cctctctgtg ctgccagcag 1620
121 tctagctgcc agccagcttg ctgcacctcc tcccaaagcc agcagggtcg ctgctgtccc 1680
122 gtctgtgtgca agcctgtgag ctgtgtgcct gtttgcctcg gggcttcttc ttcatgtctg 1740
123 cagcaatcta gctgccagcc agcttgcctg accacctctc gctgcagacc ctctctctcc 1800
124 gtgtccctcc tctgccgccc cgtgtgcagg ccgcctgct gcgtgcccg cccttctctg 1860
125 tgtgtctcca cctctcctg ccaacccagg tgcgtgccgc cagcctctcg cgtgtccctc 1920
126 ctctgacgcc ccgtgtgctc ccgcccagcc tgcgtaggcc tccgctcagg tcagaagccc 1980
129 <210> SEQ ID NO: 7
130 <211> LENGTH: 1800
131 <212> TYPE: DNA
132 <213> ORGANISM: Homo sapiens
134 <400> SEQUENCE: 7
135 g gatgagagg gggactcatg gaggaacagc cagccttga ccttgagatg gccttgcagg 60
136 gagggtaact gaaaatttac ccactgggga cagttgccta ctactaaaa cagttccagc 120
137 caccaccgca gcccttgga ggccatcccc ccagaaaatc cccaggtct cagcagggcc 180
138 ttgtccacct gtgccctcca gtgtgcacca tgtcaacctc acctaagagg ggctgtacgc 240
139 acggtcctgc aggtgcggac tctgggtcct gacagcccat gcggaacctg gtgccccag 300
140 aggagggcct ggggcagtgc cagttttggg gaatcatgtg catccatcca cccactccat 360
141 gatgctttcg tctgtatcga gtcccttgtc tccgcgcag gtgcagcagc cctcctctct 420
142 ccccccgcct gctgtctgaa cgggcagaa cctcgggcgg gcgcacaca gggagggtga 480
143 ccaggcctgg agctgtagt gcccgagccc caggccagct tcttggagg tgacctgca 540
144 ggggtgggctc tcccagggtg gacagtgggt gggacagtcc tggggcctgg agagccccac 600
145 agcccagggc acggcagcca atgaccaggc tcaggaagac ccaggcatgg aggtgtagcc 660
146 gggactgagc ctctctgggc gtggtgtgta gttccacctg gtgacccct ggaggagtta 720
147 ggccactgtc ccccgtagt tctagggtta gtcactcatt catagaaaca gtcatggcta 780
148 gagagcaatc tgagctcaaa accatgtatc cccaggagca ctacagaaaa agagaatcag 840
149 gcgaccaagg ggagtttatt ggggagcagg aggaggtgct gacaggttca agtcagggcc 900
150 aagtgacctg gggcagagaa gctgggaggg aggacagggg acccaacagg caggtgggcc 960
151 cctgtctggga ggcaggagct ggggagcttc gaggatggag attcctggga gtatggaggg 1020
152 gggggtcacc tcagcacatg ggggccccgt cccaagcggg ggcaacctcc taaccagagt 1080
153 caggaccagt tggccctggg ggtgtgcac atcagcaact ggactcctgg cctgagcaga 1140
154 ggcctcagca ggccaggcgg gacacgcgg ggcggcagag gaggggacag caggaggccg 1200
155 ggcggcagca gctggcctgg taggaggagg caggggcaca gcaggaggag atgggcacgc 1260
156 agcaggcggg cctgcatatg ggcggcaga ggaggacac ggaggaggag ggtctgcagc 1320
157 aggaggtggt gcagcaagcc ggctgacagc tagactgctg gcagcatgaa gtggaagccc 1380
158 cagagcagac gggcacacag cagatgggtt tgaagcagac aggccttcaa cagacaggca 1440
159 cgtagcagga ctgctggcag ggggaggagg tgcagcaagt cggctggcag ctagaatgct 1500
160 ggcagcatga agaggaatcc tcagaacagg tgggcacaca gcacacgggc ttgcagcaga 1560
161 caggcacaca gcaggactgc tggcaggagg aagaggcaca gcaagtggc tggcagctag 1620
162 actgctggca gcatgaagag gaatccttag agcagggtgg caggcagcac acaggcttgc 1680
163 agcagacggg cagcagcag gcctgctggc agggggaggga ggcgcagcaa gccggctggc 1740
164 agcacgaggg cgtgcaggag ctggtgcagc ctgattggca ggggctgggc tcacaggccg 1800
167 <210> SEQ ID NO: 8
168 <211> LENGTH: 1620
169 <212> TYPE: DNA
170 <213> ORGANISM: Plasmodium falciparum
172 <400> SEQUENCE: 8

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/950,051

DATE: 03/04/2002

TIME: 15:03:28

Input Set : A:\34950209.app

Output Set: N:\CRF3\03042002\I950051.raw

```

173 tatgcacgtg tgtgtgtgta taatatacaa ttcttatttc catattacct tgtctacata 60
174 tgggacattt cgtattttgc atcctaacaa catttgcaca cttcaaaata aaaaaaaaaa 120
175 aaaattatat atatatatat atatattata tatacatgta tttactttta atattcatat 180
176 tatttcgttt taaagtacca cattacaaag acacatatgt ctacaaggta aaattgctgt 240
177 atctttttca tccgtcaggc atattacaca ttcccttaaa aataaaaacat aaatataaat 300
178 aagcatataa tatatatagt atatataatg tacattcata cgtataaatg aataaaaata 360
179 tacatatata tatatatata tatatatata tgatgtacac acctgtcaga caaataatta 420
180 tttacaggat ccgtttttgt ttctggagat ttttctattc caaaaatttc ttgaacctca 480
181 aatgatctgt taccaaattg aatcttctgt ctataaagaa ttattttata tttatcttga 540
182 ttgtttttat tctcgttcat ctgattttct tgtaagtagg catagttgta ttgagcctga 600
183 agtatggcgc taccaatcgc ttttaaggact attaaaatag gaatgatata ttcataacta 660
184 ggtttacatt ttatatcatt tatattcatg ttattcaaat ttattccttc atttaattca 720
185 ctcatataaa cttgggtttaa attcttcgga aatatttttg taatcgtttg atacttttga 780
186 ggacaatatc tagaaagaaa agaaaattaa caaaacgaac ataaaataat tatttaaaaa 840
187 aaaatataaa aaaaaaaaaa aattgataaa taaacataaa atttctacat aataaatcta 900
188 atatataatta atatatatat atatatatat atatatatat tataatacat tttcttttat 960
189 tttaaaaata ttacatggtt tcttttgctt ctgttaactc ctctttacaa caaaaataaa 1020
190 tagatacctc tacatcatgt aaggcatcaa aaatgaaatt tattaataaa ttattatttc 1080
191 ccacatttat aatttttaat gtctttctac gtaagttaag tgaatttctt acaacggatg 1140
192 ttctttgaat atttatattt ggtccttgcg atattagttc acacaagttg gtgttatttg 1200
193 ttctctatca ttaagtcaaa atagaagaga aaaaaaatga aatgaaataa gaagtatact 1260
194 tacacataaa tgtacatata tatatatata ttttatattt tttatatatt ttgcttctta 1320
195 taaagtgatt acattttatta tgggattatc aaatgatatt gtttgtggct gtttggaac 1380
196 aaaaaaaaaa aaaaaaaaaa aaaagttttt aattaaatat gtgttatata taatataatg 1440
197 agaagatata tatatatata tatatggagc aaaatattat cattcttttt ttctttttta 1500
198 aaaccttaca ccatcatcat aataatcatc tcgattgtta gataacgatg aaccatttat 1560
199 tcccattata caacaagaaa attataaaat gtaacatata tatttttaag cattataaat 1620
202 <210> SEQ ID NO: 9
203 <211> LENGTH: 1560
204 <212> TYPE: DNA
205 <213> ORGANISM: Plasmodium falciparum
207 <400> SEQUENCE: 9
208 acttataaat aggttgtata caaatatata ttatttatat tttattttgt taataaattc 60
209 ttggttaacaa tttctttcca tcatcatagg gaagatatat actcttatct ttcggacgat 120
210 agcctgaaat atttaaaaaa aaataaataa atatatatat atataatata tataattacc 180
211 ttttgataat atatatatat atatatatat attttttttt tcattttacct ccatatgata 240
212 tttgctgaca caatgctggt aaacgacgat ttgatttttt ctggttataa aaatcagccc 300
213 tatataaata aagcaaaata taaaacgtat atatatatat atatattata tatattttat 360
214 tatttgtgta caagaattac catataaaat aggcctgtga aaaatgatca cttgataaat 420
215 taaatatatg tctcatttga agaatactat atttacttaa ctctgttaca attttccgc 480
216 attgcttcac ttttttttca tataaatgcc tgaaaagggg aaaatacata aataaattaa 540
217 aataataata tatatatata tatatattta ttctttttat attttttttt ttttatgtaa 600
218 actaactccc ataatatatc cttcaacaaat gatggaaaaa tatctacatt gttagccgct 660
219 tcgtctaaty ataccaaaaa ctgaaaaaag aaaataaaac attcaagtgt attatcatta 720
220 ttgtttattac atatattgat ataaataaat atatatatat atatatatat atatatatat 780
221 atatataaac aaatgatata ttctttttta ttaaaaaaaa aaaaaagctc actccaaatg 840
222 catattcttg acaattcaaa agttctgcat taaaagaata ttcaggaaca tcttttaaaa 900
223 gcatccaatc acgaacaatt ctccacgtta tatcaaaaat atatgtattt ataatatctc 960
224 cttgatactg ttcatgtctt aatcttttat taattatcca aagatgcaaa acataaaaaa 1020

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/950,051

DATE: 03/04/2002

TIME: 15:03:28

Input Set : A:\34950209.app

Output Set: N:\CRF3\03042002\I950051.raw

```
225 acattttcat attaaaagat tcattttatac gaaataagtt taatatctct tcattttcta 1080
226 atctttcaat tattaaatgt actaatctcc atgcacattc tccatattct gtatgttttc 1140
227 taaataaata ttaaaatata tatattttta ttttatatat aaacataaat tcaaaatgaa 1200
228 acttttttaa catattagac gtacatatgc tatgtttttt aaaatattat atatatatat 1260
229 tatgaagaat tgttttctta caaaaaaaat ttatatatca tagaacttat tatagaacga 1320
230 tcacattgtg gtataggat tgcatatctt gaagattctc taataactaat tttgtattc 1380
231 ttatctcat acaaaacttt actttcgata atttttaaat tatcatatac ttttatttca 1440
232 ctagaaaaat agcgatagtt aatatatctt ctctttaaaa caaattttgt acattttatac 1500
233 cacatgacaa caaaatttaa ttaaaaatta atgaaaaaaa aaaaaaaaaa aaaaaaaaaa 1560
```

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/950,051

DATE: 03/04/2002

TIME: 15:03:29

Input Set : A:\34950209.app

Output Set: N:\CRF3\03042002\I950051.raw